Automatic Classification of Primary Progressive Aphasia Patients Using Lexical and Acoustic Features

Sunghye Cho[1], Naomi Nevler[2], Sanjana Shellikeri[2], Sharon Ash[2], Mark Liberman[1], Murray Grossman[2]

Linguistic Data Consortium[1], Penn Frontotemporal Degeneration Center[2]
University of Pennsylvania, 3400 Walnut Street, Philadelphia, PA, USA

{csunghye, myl}@ldc.upenn.edu
{naomine, Sanjana.Shellikeri, ash, mgrossma}@pennmedicine.upenn.edu

Abstract

Two variants of primary progressive aphasia (PPA) are subtypes of frontotemporal degeneration (FTD), which is the most common type of dementia among individuals under 60 years of age. Semantic variant PPA (svPPA) patients present with semantic deficits in single word use, whereas nonfluent/agrammatic PPA (naPPA) patients produce simplified speech with frequent speech errors and slow speech rates. In this study, we built machine learning systems to classify PPA patients (n=63) and healthy elderly controls (n=36). We automatically extracted 18 lexical and 21 acoustic features with a natural language processing library and a speech activity detector, and we trained classifiers, experimenting with various feature selection and reduction techniques. Our models showed high accuracy, correctly distinguishing patients from controls in more than 90% of cases, svPPA patients from naPPAs in about 89% of cases, and controls, svPPA, and naPPA patients in more than 80% of cases. Our results show that classification of PPA patients using automatically derived linguistic features from digitized speech samples is highly promising, and could potentially be applied in community settings for prescreening. We plan to extend this project by including more features and additional FTD subgroups in the near future.

Keywords: Primary progressive aphasia, automatic classification, narrative speech

1. Introduction

Frontotemporal degeneration (FTD) is a type of focal dementia caused by atrophy in the brains' frontal and temporal lobes. It is the most common type of neurodegenerative disease among people under 60 years of age (Ratnavalli et al., 2002). Since individuals diagnosed with FTD are relatively young, usually still in the workforce, the personal and societal costs of the disease are substantial. For example, FTD diagnosis often results in early departure from the workforce, increasing economic burden for a household with an FTD patient and negatively affecting not only patients but also the quality of life of their families (Galvin et al., 2017). Because there are no disease-modifying drugs approved for FTD, earlier screening and slowing the apparent disease progression rate through behavioral adjustments to the environment are key to helping patients and their families. This paper proposes three machine learning systems to automatically classify two subgroups of FTD that could potentially be applied in prescreening.

About half of patients with FTD present with a linguistic impairment known as primary progressive aphasia (PPA), and sometimes this can be accompanied by a social-behavioral impairment known as behavioral variant FTD (bvFTD). There are several variants of PPA. Among these subgroups, semantic variant PPA (svPPA) patients are characterized by impaired confrontation naming, frequent substitution of pronouns for nouns, and difficulty in processing concrete words, although they show intact prosody and syntax (e.g., Amici et al., 2007; Bonner et al., 2016; Cousins et al., 2016; Nevler et al., 2019). Nonfluent/agrammatic PPA (naPPA) patients, on the other hand, present with effortful speech, slow speech rates, frequent speech errors, simplified grammar, and difficulty in retrieving verbs (e.g., Ash et al., 2009; Grossman et al., 1996; Rhee et al., 2001). Patients with either of the two subtypes have frontotemporal lobar degeneration spectrum pathology, which is commonly associated with misfolding of TDP-43 or tau proteins.

Since PPA patients show salient linguistic characteristics, we would expect automatic classification by means of linguistic features to yield high levels of accuracy. There are a few previous studies that have pursued this approach. Fraser et al. (2014) extracted 58 lexical and semantic features from the speech samples of 10 svPPA and 14 naPPA patients and 16 controls. The authors trained classifiers only with significant features for three different tasks: control versus svPPA, control versus naPPA, and svPPA versus naPPA. Their models for controls versus svPPA/naPPA showed high levels of accuracy, from 90% to 100%. However, their best performance for classifying svPPA and naPPA patients was only 79.2% accurate, suggesting that classifying patient groups is more difficult than distinguishing patients from controls. Peintner et al. (2008) extracted 41 acoustic, 81 LIWC (Language Inquiry and Word Count; Pennebaker et al., 2001), and 11 lexical features from 39 participants (9 bvFTD, 8 naPPA, 13 svPPA, and 9 controls), and trained classifiers for various classification tasks. Their composite feature set (significant features from each feature set) showed accuracy over 90% in most classification tasks, except control versus bvFTD and four-way classification. However, they did not list what features were used in the composite set, making it difficult to reproduce their results. Themistocleous et al. (2019) extracted 14 acoustic features, such as mean fundamental frequency and amplitude differences between the first and second harmonics, from 50 patients (17 logopenic variant PPA (lvPPA), 14 svPPA, 11 naPPA, and 8 naPPA with apraxia of speech) and trained classifiers with 3-fold group cross validations and a one-against-all strategy. Their models correctly identified naPPA 82% of the time and svPPA 66% of the time. The authors only used acoustic features, which explains why the accuracy of svPPA patients, who rarely show impairments in prosody, was relatively low. More importantly, all previous studies have had relatively small datasets, raising

the question of whether their results could be generalized to larger datasets. In this paper, we studied 99 participants (63 patients and 36 controls) to investigate whether lexical and acoustic features could predict the diagnostic status of the participants.

2. Objectives

Our objectives were to train three predictive models for classifying (1) controls vs. patients, (2) svPPA vs. naPPA patients, and (3) controls, svPPA and naPPA patients, experimenting with different feature selection and reduction techniques, and to identify predictive features for classifying PPA patients.

3. Methods

3.1 Participants

Our participants consisted of 63 patients diagnosed clinically with either svPPA or naPPA and 36 healthy elderly controls. Forty-two of the 63 patients had svPPA and 21 were naPPA patients. The patients were diagnosed by experienced neurologists at the Department of Neurology of the Hospital of the University of Pennsylvania in accordance with published criteria (Gorno-Tempini et al., 2011). Of the 42 svPPA patients, 32 showed concomitant mild behavioral symptoms, which is a common co-occurrence in this group. We focused on frontotemporal lobar degeneration (FTLD) spectrum pathology in this study, and so we did not include lvPPA patients, who most often have Alzheimer's pathology. Our participants were matched on sex ratio and education levels, but not on age, because naPPA patients on average have an later disease onset than svPPA patients (Johnson et al., 2005). The patient groups did not differ on the Mini Mental State Exam scores (MMSE) or disease durations, but they significantly differed on the Boston Naming Test (BNT) scores, which is expected due to svPPA patients' difficulty in naming tasks. All participants were native speakers of English. The study was approved by the Institutional Review Board of the Hospital of the University of Pennsylvania, and all participants signed a written consent form. Participants' demographic and neuropsychological characteristics are summarized in Table 1.

	controls	**svPPA**	**naPPA**	***p*-value**
Age	68.5 (7.9)	63.3 (6.9)	70.4 (9.4)	0.001
Sex	23 F/13 M	23 F/19 M	11 F/11 M	0.483
Education (years)	15.9 (2.5)	15.1 (2.8)	15.3 (3.1)	0.408
MMSE (range: 0-30)	29.2 (1)	22.1 (6.3)	22.7 (5.9)	< 0.001
BNT (range: 0-30)	27.9 (2.5)	7.5 (6.4)	24.7 (4.6)	< 0.001
Disease duration (yrs)	NA	3.9 (2)	3.2 (1.9)	0.214
Total number of words in Cookie Theft	174.4 (66.4)	148.1 (62.8)	91.0 (55.8)	< 0.001

Table 1: Mean (SD) demographic and neuropsychological characteristics of the participants. MMSE: Mini Mental State Exam, BNT: Boston Naming Test.

3.2 Data

The Cookie Theft picture from the Boston Diagnostic Aphasia Examination (Goodglass et al., 1983) was used to elicit narrative speech from the participants. Participants described the picture for about one minute, and their descriptions were digitally recorded. Some patients made several recordings, but we used the earliest recording of each participant in this analysis in order to differentiate among the conditions early in the disease course. An expert linguist generated verbatim transcription of the picture descriptions, including all non-verbal speech, hesitations and false starts, and a team of trained annotators at the Linguistic Data Consortium (LDC) of the University of Pennsylvania reviewed and revised the annotations for quality checking.

4. Feature Extraction

4.1 Lexical Features

We ran a POS tagger in spaCy (Honnibal & Johnson, 2015) to automatically tag POS categories of all words that the participants produced in the picture descriptions. Before running the tagger, we cleaned the transcripts by removing interviewers' prompts and annotations for non-verbal speech. A professional linguist manually validated the accuracy of spaCy with a subset of our data (n=21). The mean group accuracy varied from 95% (controls) to 90% (PPAs). There was no significant difference in the accuracy among patient groups (p>0.05). Since the accuracy of the spaCy POS tagger with their basic model ('en_core_web_sm') was high, we did not train a POS tagger separately in this study. The POS tokens were tallied per participant, and the count of each POS category per 100 words was calculated (= (raw counts/total number of words) * 100). In addition to the frequency of each POS category, we measured the number of tense-inflected verbs and unique nouns per 100 words. We summed the number of modal auxiliary verbs, past tense verbs and present tense verbs that spaCy tagged to count the number of tense-inflected verbs per 100 words. The number of noun lemmas was used for the number of unique nouns per 100 words.

We also rated nouns that participants produced for concreteness (Brysbaert et al., 2014), semantic ambiguity (Hoffman et al., 2013), word frequency (Brysbaert & New, 2009), age of acquisition (AoA; Brysbaert et al., 2018) and word familiarity (Brysbaert et al., 2018) for their potential to distinguish svPPA patients from others. Since the published norms we used had a limited number of words, we rated the lemma of a noun if a noun itself was not listed in the published norms. A noun was not rated if neither the noun nor its lemma was listed in the norms. In total, we had 18 text-related features: POS counts per 100 words (nouns, verbs, adjectives, adverbs, prepositions, determiners, conjunctions, interjections, pronouns, and speech errors/partial words—[X] in spaCy), number of tense-inflected verbs and unique nouns per 100 words, lexical features of nouns (concreteness, ambiguity, frequency, AoA, familiarity), and total number of words.

4.2 Acoustic Features

We used an in-house Gaussian Mixture Models-Hidden Markov Models based Speech Activity Detector (SAD) developed at the LDC to segment the recordings into speech and silent pause segments. We set the minimum duration of a speech segment at 250 ms and that of a silence segment at 150 ms. We reviewed the outputs of SAD, corrected wrong segmentations, and excluded interviewers' speech and non-verbal speech segments. Using the durations of speech and silent pause segments, we extracted 12 durational features:

- The mean duration of speech and pause segments
- The number of total pauses and speech segments
- Total speech time (speech only)
- Total pause time (pause only)
- Total time (speech time + pause time)
- Sample duration (duration of the entire recording)
- Percent of speech time of the total time
- Breath frequency (= number of pauses over total time)
- Speech frequency (= number of speech segments over total time)
- Pause rate per minute (= number of pauses over total speech time)

We also pitch-tracked speech segments of the participants with a script in Praat (Boersma & Weenink, 2020) and extracted the 10^{th} to 90^{th} fundamental frequency (f0) percentile values for each speaker. To minimize individual differences in pitch due to physiological factors, such as sex, height, and the size of the larynx, the extracted f0 values in Hz were converted to semitones (St) using each speaker's 10^{th} percentile as a baseline: $St = \log_2(Hz / 10^{th} \text{ percentile})*12$. We had 21 acoustic features in total, including pitch percentile values along with the 12 durational features. The final feature set included 18 lexical and 21 acoustic features and 3 demographic characteristics of the participants: age, sex, and education level.

5. Model Training

We trained two different machine learning algorithms from the scikit-learn package (Pedregosa et al., 2011) in Python: Random Forest and Support Vector Machine (SVM) classifiers. In all models, we imputed missing values with a mean value using SimpleImputer and standardized features with StandardScaler in scikit-learn for effective learning. We performed leave-one-out cross-validation (CV) to evaluate the generalizability of the models and reported the average accuracy of all CV folds.

We experimented with feature selection and reduction methods. For feature selection, we performed t-tests (for binary classifications) and trained models with features that were significant at the level of $p < 0.05$, 0.01, 0.005, and 0.001. We used the same feature set used in the control-patient pairwise classification for the three-way classification (control vs. svPPA vs. naPPA). We compared the performance of models trained with selected features and a model without any feature selection. For feature reduction, we performed Principal Component Analysis (PCA) and trained models, varying the number of components from 1 to 10. We compared the performance of models trained with PCA components and that of a model trained without any feature reduction and reported the best performance after tuning hyperparameters.

6. Classification Results

6.1 Binary Classification between Controls and Patients

An SVM classifier trained with all features which were reduced to 10 PCA components performed best in this classification task, showing 90.9% accuracy and 0.94 area under the curve (AUC). Our model correctly identified 33 controls out of 36 and 57 patients out of 63. The full classification report is shown in Table 2, and the receiver operating characteristic (ROC) curve for this contrast is provided in Figure 1.

	Accuracy	Precision	Recall	F1-score
Controls	0.92	0.85	0.92	0.88
Patients	0.90	0.95	0.90	0.93
Weighted average	0.91	0.91	0.91	0.91

Table 2: Classification report of the SVM classifier for the classification of patients and controls.

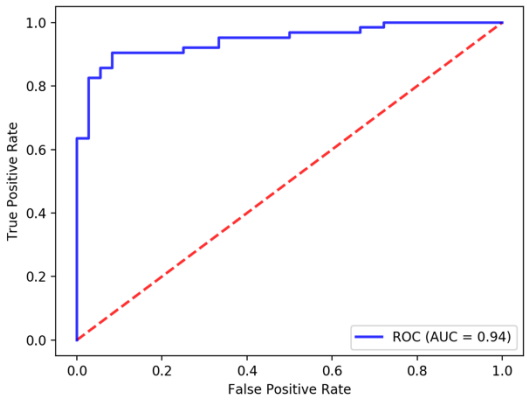

Figure 1: Receiver Operator Characteristic Curve for the classification of controls and patients.

6.2 Binary Classification of Patient groups

A Random Forest classifier trained with features that were significant at the level of $p<0.005$ and reduced to three PCA components performed best in this classification task. The model showed 88.9% accuracy with 0.87 AUC. The model correctly identified 40 svPPA patients out of 42 and 16 naPPA patients out of 21. Our model resulted in a higher F1-score for classifying svPPA patients (0.92) than naPPA patients (0.82), suggesting that in general identifying naPPA patients was more difficult than identifying svPPA patients. The full classification scores are in Table 3, and the ROC curve for this contrast is provided in Figure 2.

	Accuracy	Precision	Recall	F1-score
svPPA	0.95	0.89	0.95	0.92
naPPA	0.76	0.89	0.76	0.82
Weighted average	0.89	0.89	0.89	0.89

Table 3: Classification report of the Random Forest classifier for the classification of svPPA and naPPA patients.

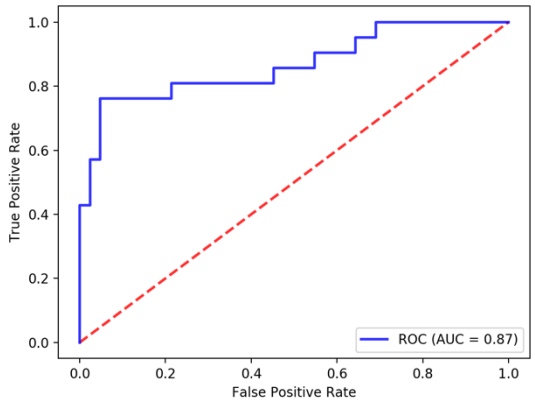

Figure 2: Receiver Operator Characteristic Curve for the classification of svPPA and naPPA patients.

The features that were selected included counts of nouns, pronouns, verbs, tense-inflected verbs, speech errors/partial words, unique nouns per 100 words; concreteness, semantic ambiguity, frequency of nouns; participants' age and total number of pauses. Figure 3 shows group differences in the selected features.

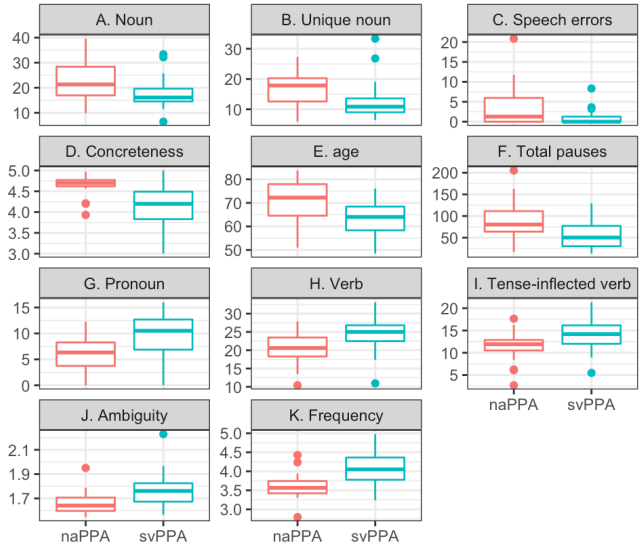

Figure 3: Group differences in selected features for the classification of svPPA and naPPA patients. The POS counts and the numbers of tense-inflected verbs and unique nouns are per 100 words. The top two rows show features where values of naPPA patients are significantly higher than those of svPPA and the bottom two rows show features where values of svPPA patients are significantly higher than those of naPPA (both at $p<0.005$).

Among the 11 selected features, most were lexical, and only one acoustic feature, total number of pauses, was selected. As expected, semantic aspects of nouns that patients produced, such as concreteness and semantic ambiguity, were important features in distinguishing svPPA patients from naPPA patients. Further discussion of the acoustic features in PPA patients can be found in Nevler et al. (2019), and further discussion of the lexical features can be found in Cho et al. (under review).

6.3 Three-way Classification

An SVM classifier trained with all features without any feature reduction performed best for the three-way classification, yielding 80.8% accuracy with 0.9 macro-averaged AUC. The model correctly identified 32 controls out of 36, 34 svPPA patients out of 42, and 14 naPPA patients out of 21. The model's F1-score is high for controls and svPPA patients (> 0.8), but it was below 0.7 for naPPA patients, again suggesting that naPPA patients were difficult to identify. The full classification report and the confusion matrix are provided in Tables 4 and 5, and the ROC curve for this contrast is provided in Figure 4.

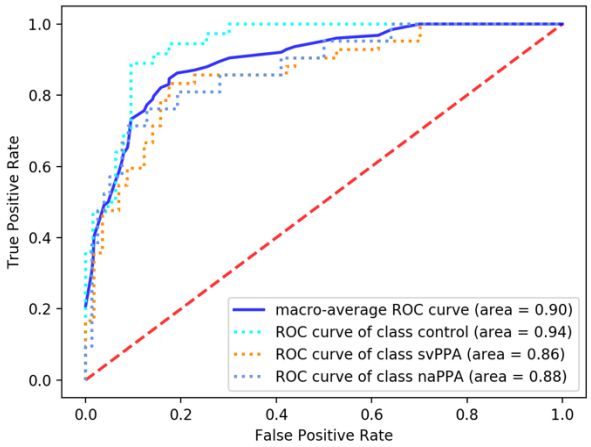

Figure 4: Receiver Operator Characteristic Curve for the classification of controls and svPPA and naPPA patients.

	Accuracy	Precision	Recall	F1-score
Control	0.89	0.84	0.89	0.86
svPPA	0.81	0.83	0.81	0.82
naPPA	0.67	0.70	0.67	0.68
Weighted average	0.81	0.81	0.81	0.81

Table 4: Classification report of the SVM classifier for the three-way classification.

	Control	svPPA	naPPA
Controls	32	2	2
svPPA	4	34	4
naPPA	2	5	14

Table 5: Confusion matrix of the three-way classification (column: actual, row: predicted). The number of accurately classified participants is highlighted in gray.

7. Discussion and Conclusion

This paper reports the results of automatic classification systems for three classification tasks: i) control versus patients, ii) svPPA versus naPPA patients, and iii) control versus svPPA versus naPPA. We automatically extracted 18 lexical features from one-minute narrative speech samples using spaCy, one of the most modern, state-of-the-art natural language processing libraries in Python. We also automatically extracted 21 acoustic and durational features with SAD. Using these features with additional demographic information, we trained three machine learning classifiers, experimenting with different feature selection and reduction techniques, and used leave-one-out cross-validation. We found group differences in the selected features. Our model for the control versus patient classification trained with all features, which were reduced to 10 PCA components, correctly distinguished patients from controls in more than 90% of cases. Our classifier for the svPPA versus naPPA task selected 11 features (9 lexical, 1 acoustic and 1 demographic), which were later reduced to 3 PCA components. Our classifier correctly identified the diagnostic group of the patients with 88.9% accuracy, which outperformed the system for the same task in previous studies (79.2% in Fraser et al., 2014; 82% for naPPA patients in Themistocleous et al., 2019). Lastly, our system for the three-way classification, which was trained with all features without any feature reduction, showed high overall accuracy (over 80%) in classifying controls, svPPA and naPPA patients, which is much higher than the chance level (33.3%). The performance of the systems in this report is highly promising in that we only had one-minute narrative speech samples, which are quick and easy to collect. We believe that this line of research could potentially benefit populations with the earliest features of PPA.

Our models performed well, but there is still room for improvement, in particular, for the three-way classification system, where classification of naPPA was < 80%. In the future, we plan to include more features, such as letter or category fluency scores, Mel-frequency cepstral coefficients, or word frequency as log-odds ratio (Monroe et al., 2008) to improve the performance of the models. We also aim to extend our research by including more patient groups. First, we would consider evaluating patients with lvPPA, which is another variant of PPA associated with Alzheimer's disease pathology, with frequent filler words (*um* or *uh*) as a prominent feature. Second, we would consider including bvFTD patients, who have pathology similar to that of svPPA and naPPA patients. Although without obvious aphasia, these patients do have subtle speech deficits (Nevler et al., 2018). In addition, we plan to collect conversational data in the near future to explore subtle group differences among these patient groups that may not have been captured in monologue, narrative speech samples. In natural conversation, speakers employ a variety of prosodic features to deliver the intended message effectively. We believe these additional features will improve the models' performance.

8. Acknowledgements

We thank the participants and their family members for participating in the study and the research assistants who helped collect the data. This study was funded by National Institutes of Health (AG017586, AG053940, AG052943, NS088341, DC013063, AG054519), the Institute on Aging at the University of Pennsylvania, the Alzheimer's Association (AACSF-18-567131), an anonymous donor, and the Wyncote Foundation.

9. Bibliographical References

Amici, S., Ogar, J., Bozeat, S., Arnold, R., Watson, P., and Hodges, J.R. (2006). Performance in specific language tasks correlates with regional volume changes in progressive aphasia. *Cognitive and Behavioral Neurology,* 20(4): 203–211.

Ash, S., Moore, P., Vesely, L., Gunawardena, D., McMillan, C., Anderson, C., … and Grossman, M. Non-fluent speech in frontotemporal lobar degeneration. *Journal of Neurolinguistics,* 22(4):370–383.

Boersma, P., and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program]. Version 6.1.09, retrieved 26 January 2020 from http://www.praat.org

Bonner, M., Price, A., Peelle, J., and Grossman, M. (2016). Semantics of the visual environment encoded in parahippocampal cortex. *Journal of Cognitive Neuroscience,* 28(3): 361–378.

Brysbaert, M., and New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4):977–990.

Brysbaert, M., Mandera, P., and Keuleers, E. (2018). Word prevalence norms for 62 ,000 English lemmas. *Behavior Research Methods*, 51(2):467–479.

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3):904–911.

Cho, S., Nevler, N., Ash, S., Shellikeri, S., Irwin, D., Massimo, L., Rascovsky, K., Olm, C., Grossman, M., and Liberman, M. (under review). Automated analysis of lexical features in Frontotemporal Degeneration.

Cousins, K., York, C., Bauer, L., and Grossman, M. (2016). Cognitive and anatomic double dissociation in the representation of concrete and abstract words in semantic variant and behavioral variant frontotemporal degeneration. *Neuropsychologia,* 84:244–251.

Fraser, K., Meltzer, J., Graham, H., Leonard, C., Hirst, G., Black, S., and Rochon, E. (2014). Automated

classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex,* 55:43–60.

Galvin, J., Howard, D., Denny, S., Dickinson, S., and Tatton, N. (2017). The social and economic burden of frontotemporal degeneration. *Neurology,* 89:2049–2056.

Goodglass, H., Kaplan, E., and Weintraub, S. (1983). Boston diagnostic aphasia examination. Philadelphia, PA, Lea & Febiger.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology,* 76:1006–1014.

Grossman, M., Mickanin, J., Onishi, K., Hughes, E., D'Esposito, M., Ding, X. S., … and Reivich, M. (1996). Progressive nonfluent aphasia: Language, cognitive, and PET measures contrasted with probable Alzheimer's disease. *Journal of Cognitive Neuroscience*, 8(2) :135–154.

Honnibal, M., and Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In Màrques et al., Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 1373–1378, Lisbon, Portugal, September.

Johnson, J.K., Diehl, J., Mendez, M.F., Neuhaus, J., Shapira, J.S., Forman, M., Chute, D.J., Roberson, E.D., Pace-Savitsky, C., Neumann, M., Chow T.W., Rosen, H.J., Forstl, H., Kurz, A., and Miller, B.L. Frototemporal lobar degeneration: Demographic characteristics of 353 patients. *Arch Neurol.,* 62: 925–930.

Monroe, B., Colaresi, M., and Quinn, K. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16:372–403.

Nevler, N., Ash, S., Irwin, D. J., Liberman, M., and Grossman, M. (2019). Validated automatic speech biomarkers in primary progressive aphasia. *Annals of Clinical and Translational Neurology,* 6(1):4–14.

Nevler, N., Ash, S., Jester, C., Irwin D. J., Liberman, M., and Grossman, M. (2018). Automatic measurement of prosody in behavioral variant FTD. *Neurology,* 89:650–656.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, B., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research,* 12:2825–2830.

Peintner, B., Jarrold, W., Vergyri, D., Richey, C., Gorno-Tempini, M., and Ogar, J. (2008). Learning diagnostic models using speech and language measures. In Wheeler B., et al., Proceedings of the 30th Annual International IEEE EMBS Conference, pages 4648–4651, Vancouver, Canada, August.

Ratnavalli, E., Brayne, C., Dawson, K., and Hodges, J.R. (2002). The prevalence of frontotemporal dementia. *Neurology,* 58:1615–1621.

Rhee, J., Antiquena, P., and Grossman, M. (2001). Verb comprehension in frontotemporal degeneration: The role of grammatical, semantic and executive components. *Neurocase*, 7(2):173–184.

Themistocleous, C., Ficek, B., Webster K.T., Wendt H., Hillis A.E., Den Ouden, D.B., and Tsapkini, K. Acoustic markers of PPA variants using machine learning. *Frontiers in Human Neuroscience 92.* doi: 10.3389/conf.fnhum.2018.228.00092.